

## **RAW SEQUENCE LISTING**

**The Biotechnology Systems Branch of the Scientific and Technical  
Information Center (STIC) no errors detected.**

Application Serial Number: 09/858,332G  
Source: IFW/6  
Date Processed by STIC: 07-18-2005

# ***ENTERED***



IFW16

## RAW SEQUENCE LISTING

DATE: 07/18/2005

PATENT APPLICATION: US/09/858,332G

TIME: 14:00:13

Input Set : D:\seqlist.txt

Output Set: N:\CRF4\07152005\I858332G.raw

4 <110> APPLICANT: Tchaga, Grigory S.  
 5       Jokhadze, George  
 7 <120> TITLE OF INVENTION: Metal Ion Affinity Tags and Methods for  
 8       Using the Same  
 11 <130> FILE REFERENCE: CLON-056CIP  
 13 <140> CURRENT APPLICATION NUMBER: US 09/858,332G  
 14 <141> CURRENT FILING DATE: 2001-05-15  
 16 <150> PRIOR APPLICATION NUMBER: 09/404,017  
 17 <151> PRIOR FILING DATE: 1999-09-23  
 19 <150> PRIOR APPLICATION NUMBER: 60/101,867  
 20 <151> PRIOR FILING DATE: 1998-09-25  
 22 <160> NUMBER OF SEQ ID NOS: 34  
 24 <170> SOFTWARE: FastSEQ for Windows Version 4.0  
 26 <210> SEQ ID NO: 1  
 27 <211> LENGTH: 16  
 28 <212> TYPE: PRT  
 29 <213> ORGANISM: Artificial Sequence  
 31 <220> FEATURE:  
 32 <223> OTHER INFORMATION: affinity peptide  
 34 <400> SEQUENCE: 1  
 35 His Leu Ile His Asn Val His Lys Glu Glu His Ala His Ala His Asn  
 36   1                   5                   10                   15  
 39 <210> SEQ ID NO: 2  
 40 <211> LENGTH: 18  
 41 <212> TYPE: PRT  
 42 <213> ORGANISM: Artificial Sequence  
 44 <220> FEATURE:  
 45 <223> OTHER INFORMATION: affinity peptide  
 47 <400> SEQUENCE: 2  
 48 His Asp Asp His Asp Asp His Asp Asp His Asp Asp His Asp Asp His  
 49   1                   5                   10                   15  
 50 Asp Asp  
 54 <210> SEQ ID NO: 3  
 55 <211> LENGTH: 18  
 56 <212> TYPE: PRT  
 57 <213> ORGANISM: Artificial Sequence  
 59 <220> FEATURE:  
 60 <223> OTHER INFORMATION: affinity peptide  
 62 <400> SEQUENCE: 3  
 63 His Glu Glu His Glu Glu His Glu Glu His Glu Glu His  
 64   1                   5                   10                   15  
 65 Glu Glu  
 69 <210> SEQ ID NO: 4

(pg-6)

RAW SEQUENCE LISTING                      DATE: 07/18/2005  
 PATENT APPLICATION: US/09/858,332G        TIME: 14:00:13

Input Set : D:\seqlist.txt  
 Output Set: N:\CRF4\07152005\I858332G.raw

```

70 <211> LENGTH: 18
71 <212> TYPE: PRT
72 <213> ORGANISM: Artificial Sequence
74 <220> FEATURE:
75 <223> OTHER INFORMATION: affinity peptide
77 <400> SEQUENCE: 4
78 His Asp Glu His Asp Glu His Glu Asn His Glu Asn His Glu Asp His
79 1          5          10          15
80 Glu Asp
84 <210> SEQ ID NO: 5
85 <211> LENGTH: 18
86 <212> TYPE: PRT
87 <213> ORGANISM: Artificial Sequence
89 <220> FEATURE:
90 <223> OTHER INFORMATION: affinity peptide
92 <400> SEQUENCE: 5
93 His Glu Asp His Glu Asp His Glu Asp His Glu Asp His Glu Asp His
94 1          5          10          15
95 Glu Asp
99 <210> SEQ ID NO: 6
100 <211> LENGTH: 5
101 <212> TYPE: PRT
102 <213> ORGANISM: Artificial Sequence
104 <220> FEATURE:
105 <223> OTHER INFORMATION: affinity peptide
107 <400> SEQUENCE: 6
108 Asp Asp Asp Asp Lys
109 1          5
112 <210> SEQ ID NO: 7
113 <211> LENGTH: 4
114 <212> TYPE: PRT
115 <213> ORGANISM: Artificial Sequence
117 <220> FEATURE:
118 <223> OTHER INFORMATION: enterokinase cleavage site
120 <400> SEQUENCE: 7
121 Ile Glu Gly Arg
122 1
125 <210> SEQ ID NO: 8
126 <211> LENGTH: 6
127 <212> TYPE: PRT
128 <213> ORGANISM: Artificial Sequence
130 <220> FEATURE:
131 <223> OTHER INFORMATION: a factor Xa cleavage site
133 <400> SEQUENCE: 8
134 Leu Val Pro Arg Gly Ser
135 1          5
138 <210> SEQ ID NO: 9
139 <211> LENGTH: 8
140 <212> TYPE: PRT

```

## RAW SEQUENCE LISTING

DATE: 07/18/2005

PATENT APPLICATION: US/09/858,332G

TIME: 14:00:13

Input Set : D:\seqlist.txt

Output Set: N:\CRF4\07152005\I858332G.raw

```

141 <213> ORGANISM: Artificial Sequence
143 <220> FEATURE:
144 <223> OTHER INFORMATION: a thrombin cleavage site
146 <400> SEQUENCE: 9
147 His Pro Phe His Leu Val Ile His
148 1 5
151 <210> SEQ ID NO: 10
152 <211> LENGTH: 10
153 <212> TYPE: PRT
154 <213> ORGANISM: Artificial Sequence
156 <220> FEATURE:
157 <223> OTHER INFORMATION: a renin cleavage site
159 <400> SEQUENCE: 10
160 Cys Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
161 1 5 10
164 <210> SEQ ID NO: 11
165 <211> LENGTH: 8
166 <212> TYPE: PRT
167 <213> ORGANISM: Artificial Sequence
169 <220> FEATURE:
170 <223> OTHER INFORMATION: an immunological tag
172 <400> SEQUENCE: 11
173 Asp Tyr Lys Asp Asp Asp Asp Lys
174 1 5
177 <210> SEQ ID NO: 12
178 <211> LENGTH: 11
179 <212> TYPE: PRT
180 <213> ORGANISM: Artificial Sequence
182 <220> FEATURE:
183 <223> OTHER INFORMATION: an immunological tag
185 <400> SEQUENCE: 12
186 Cys Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
187 1 5 10
190 <210> SEQ ID NO: 13
191 <211> LENGTH: 3426
192 <212> TYPE: DNA
193 <213> ORGANISM: Artificial Sequence
195 <220> FEATURE:
196 <223> OTHER INFORMATION: DNA sequence of vector containing cDNA of
197 recombinant enterokinase
199 <400> SEQUENCE: 13
200 gacgaaaggg cctcgtgata cgcctatttt tatagggttaa tgtcatgata ataatggttt 60
201 cttagacgtc aggtggcact ttctcggggaa atgtgcgcgg aaccttatt tgtttatttt 120
202 tctaaataca ttcaaataatg tatccgctca tgagacaata accctgataa atgcttcaat 180
203 aatattgaaa aaggaagagt atgagtattc aacatttccg tgcgcctt attccctttt 240
204 ttgcggcatt ttgccttcct gtttttgctc acccagaaac gctggtgaaa gtaaaagatg 300
205 ctgaagatca gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga 360
206 tccttgagag ttttcgcccc gaagaacggt ttccaatgat gagcactttt aaagttctgc 420
207 tatgtggcgc ggtattatcc cgtattgacg ccgggcaaga gcaactcggt cgccgcatac 480

```

## RAW SEQUENCE LISTING

DATE: 07/18/2005

PATENT APPLICATION: US/09/858,332G

TIME: 14:00:13

Input Set : D:\seqlist.txt

Output Set: N:\CRF4\07152005\I858332G.raw

```

208 actatttctca gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg 540
209 gcatgacagt aagagaatta tgcagtgtcg ccataaccat gagtataac actgcggcca 600
210 acttacttct gacaacgata ggaggaccga aggagctaac cgcttttttg cacaacatgg 660
211 gggatcatgt aactcgctt gatcggtggg aaccggagct gaatgaagcc ataccaaacg 720
212 acgagcgtga caccacgatg cctgtagcaa tggcaacaac gttgcgcaaa ctattaactg 780
213 gcgaactact tactctagct tcccggcaac aattaataga ctggatggag gcggataaag 840
214 ttgcaggacc acttctgcgc tgggcccttc cggctggctg gtttattgct gataaatctg 900
215 gagccggtga gcgtgggtct cgcggtatca ttgcagcact ggggccagat ggtaagccct 960
216 cccgtatcgt agttatctac acgacgggga gtcaggcaac tatggatgaa cgaaatagac 1020
217 agatcgctga gataggtgcc tactgatta agcattggta actgtcagac caagtttact 1080
218 catatatact ttagattgat ttaaaacttc atttttaatt taaaaggatc taggtgaaga 1140
219 tcctttttga taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagggg 1200
220 cagaccccgat agaaaagatc aaaggatctt cttgagatcc ttttttctg cgcgtaactc 1260
221 gctgcttgca aacaaaaaaa ccaccgctac cagcgggtgg ttggttgccg gatcaagagc 1320
222 taccactct tttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc 1380
223 ttctagtgtg gccgtagtta gcccaccact tcaagaactc tgtagcaccg cctacatacc 1440
224 tcgctctgct aatcctgtta ccagtggctg ctgccagtg cgataagtcg tgtcttaccg 1500
225 ggttggtact aagacgatag ttaccggata aggcgcagcg gtcgggctga acgggggggt 1560
226 cgtgcacaca gccagcttg gagcgaacga cctacaccga actgagatac ctacagcgtg 1620
227 agctatgaga aagcgccacg cttcccgaag ggagaaaggc ggacaggtat ccggtaaagc 1680
228 gcagggcgg aacaggagag cgcacgaggg agcttccagg gggaaacgcc tggatatctt 1740
229 atagtcctgt cgggtttcgc cactctgac ttgagcgtcg atttttgtga tgctcgtcag 1800
230 gggggcggag cctatggaaa aacgccagca acgcggcctt ttacgggttc ctggcctttt 1860
231 gctggccttt tgctcacatg ttctttcctg cgttatcccc tgattctgtg gataaccgta 1920
232 ttaccgcctt tgagtgaagt gataccgctc gccgcagccg aacgaccgag cgcagcagat 1980
233 cagtgaacga ggaagcggaa gagcgcccaa tacgcaaacc gcctctcccc gcgcgttggc 2040
234 cgattcatta atgcagctgg cacgacaggt ttcccgaact gaaagcgggc agtgagcgca 2100
235 acgcaattaa tgtgagttag ctactcatt aggcacccca ggctttacac tttatgcttc 2160
236 cggctcgtat gttgtgtgga attgtgagcg gataacaatt tcacacagga aacagctatg 2220
237 accatgatta cgccaagctt gaaggatcat ctcatccaca atgtccaca agaggagcag 2280
238 gctcatgccc acaacaagat cgatattgtc ggagggaagt actccagaga aggagcctgg 2340
239 ccttgggctg ttgctctgta ttccgacgat caacaggtct gcggagcttc tctggtgagc 2400
240 agggattggc tgggtgcggc cgcccactgc gtgtacggga gaaatatgga gccgtctaag 2460
241 tggaaaagcag tgctaggcct gcatatggca tcaaatctga cttctcctca gatagaaact 2520
242 aggttgattg accaaattgt cataaaccga cactacaata aacggagaaa gaacaatgac 2580
243 attgccatga tgcatttga aatgaaagt aactacacag attatataca gcctatttgt 2640
244 ttaccagaag aaaatcaagt tttccccca ggaagaattt gttctattgc tggctggggg 2700
245 gcacttatat atcaaggctt tactgcagac gtactgcaag aagctgacgt tccccttcta 2760
246 tcaaatgaga aatgtcaaca acagatgccga gaataataca ttacggaaaa tatggtgtgt 2820
247 gcaggctatg aagcaggagg ggtagattct tgtcaggggg attcaggcgg accactcatg 2880
248 tgccaagaaa acaacagatg gtcctgtggt ggcgtgacgt catttgata tcaatgtgca 2940
249 ctgcctaata gcccaggggt gtatgcccg gtcccaagg tccacagatg gatacaaagt 3000
250 tttctacatg agctcgtaat tagctgagaa ttactggcc gtcgttttac aacgtcgtga 3060
251 ctgggaaaac cctggcgtaa cccaacttaa tcgccttgca gcacatcccc ctttcgccag 3120
252 ctggcgtaat agcgaagagg cccgcaccga tcgccttcc caacagttgc gcagcctgaa 3180
253 tggcgaatgg cgctgatgc ggtattttct ccttacgcat ctgtgcggta tttcacaccg 3240
254 cactgggtgc actctcagta caatctgctc tgatgccgca tagttaagcc agccccgaca 3300
255 cccgccaca cccgtgacg cgcctgacg ggcttgtctg ctcccggcat ccgcttacag 3360
256 acaagctgtg accgtctccg ggagctgcat gtgtcagagg ttttcaccgt catcaccgaa 3420

```

## RAW SEQUENCE LISTING

DATE: 07/18/2005

PATENT APPLICATION: US/09/858,332G

TIME: 14:00:13

Input Set : D:\seqlist.txt

Output Set: N:\CRF4\07152005\I858332G.raw

```

257 acgcgc
259 <210> SEQ ID NO: 14
260 <211> LENGTH: 269
261 <212> TYPE: PRT
262 <213> ORGANISM: Artificial Sequence
264 <220> FEATURE:
265 <223> OTHER INFORMATION: protein sequence of vector containing cDNA of
266 recombinant enterokinase
268 <400> SEQUENCE: 14
269 Met Thr Met Ile Thr Pro Ser Leu Lys Asp His Leu Ile His Asn Val
270 1 5 10 15
271 His Lys Glu Glu His Ala His Ala His Asn Lys Ile Asp Ile Val Gly
272 20 25 30
273 Gly Ser Asp Ser Arg Glu Gly Ala Trp Pro Trp Val Val Ala Leu Tyr
274 35 40 45
275 Phe Asp Asp Gln Gln Val Cys Gly Ala Ser Leu Val Ser Arg Asp Trp
276 50 55 60
277 Leu Val Ser Ala Ala His Cys Val Tyr Gly Arg Asn Met Glu Pro Ser
278 65 70 75 80
279 Lys Trp Lys Ala Val Leu Gly Leu His Met Ala Ser Asn Leu Thr Ser
280 85 90 95
281 Pro Gln Ile Glu Thr Arg Leu Ile Asp Gln Ile Val Ile Asn Pro His
282 100 105 110
283 Tyr Asn Lys Arg Arg Lys Asn Asn Asp Ile Ala Met Met His Leu Glu
284 115 120 125
285 Met Lys Val Asn Tyr Thr Asp Tyr Ile Gln Pro Ile Cys Leu Pro Glu
286 130 135 140
287 Glu Asn Gln Val Phe Pro Gly Arg Ile Cys Ser Ile Ala Gly Trp
288 145 150 155 160
289 Gly Ala Leu Ile Tyr Gln Gly Ser Thr Ala Asp Val Leu Gln Glu Ala
290 165 170 175
291 Asp Val Pro Leu Leu Ser Asn Glu Lys Cys Gln Gln Gln Met Pro Glu
292 180 185 190
293 Tyr Asn Ile Thr Glu Asn Met Val Cys Ala Gly Tyr Glu Ala Gly Gly
294 195 200 205
295 Val Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Met Cys Gln Glu
296 210 215 220
297 Asn Asn Arg Trp Leu Leu Ala Gly Val Thr Ser Phe Gly Tyr Gln Cys
298 225 230 235 240
299 Ala Leu Pro Asn Arg Pro Gly Val Tyr Ala Arg Val Pro Arg Phe Thr
300 245 250 255
301 Glu Trp Ile Gln Ser Phe Leu His Glu Leu Val Ile Ser
302 260 265
305 <210> SEQ ID NO: 15
306 <211> LENGTH: 12
307 <212> TYPE: PRT
308 <213> ORGANISM: Artificial Sequence
310 <220> FEATURE:
311 <223> OTHER INFORMATION: an amino acid sequence embodiment of the affinity

```

RAW SEQUENCE LISTING ERROR SUMMARY      DATE: 07/18/2005  
PATENT APPLICATION:    US/09/858,332G      TIME: 14:00:14

Input Set : D:\seqlist.txt  
Output Set: N:\CRF4\07152005\I858332G.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:21; Xaa Pos. 2,3,5,6,7,9

Seq#:23; Xaa Pos. 2,3,5,6,8,9

**VERIFICATION SUMMARY**

DATE: 07/18/2005

PATENT APPLICATION: US/09/858,332G

TIME: 14:00:14

Input Set : D:\seqlist.txt

Output Set: N:\CRF4\07152005\I858332G.raw

L:421 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:21 after pos.:0

L:473 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:23 after pos.:0